CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Elsevier - Publisher Connector

On the theory of folding kinetics for short proteins

Vijay S Pande[1,2], Alexander Yu Grosberg[2,3] and Toyoichi Tanaka[2]

Background: Recent data have suggested two principles that are central to the work we describe here. First, proteins are the result of evolutionary 'sequence selection' to optimize the energy of the native state. Second, the overlap with the native state is a qualitatively suitable reaction coordinate for modeling folding kinetics. The former principle is bolder and better established.

Results: Employing only these two principles, we have constructed a non-phenomenological, correlated energy landscape theory that predicts single barrier protein folding kinetics. Moreover, we are able to analytically describe the nature of the free energetic barrier between the denatured and native states of a protein and to detail the nature of folding kinetics for short proteins. Our model predicts Hammond behavior and also describes how mutations can lead to drastic differences in folding times.

Conclusions: We find that folding and unfolding kinetics can be characterized by a single thermodynamic parameter and, moreover, that Monte Carlo simulation data on folding and unfolding rates with different temperatures and mutations collapse with this characterization. Our results also delineate a regime in which kinetics may proceed via a single unique nucleus.

Addresses: [1]Department of Physics, University of California at Berkeley, Berkeley, California 94720, USA. [2]Department of Physics and Center for Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Institute of Chemical Physics, Russian Academy of Sciences, Moscow 117977, Russia.

Correspondence: Alexander Yu Grosberg
e-mail: shura@gels.mit.edu

Key words: folding/unfolding kinetics, correlated energy landscape

Received: 27 Aug 1996
Revisions requested: 07 Oct 1996
Revisions received: 18 Dec 1996
Accepted: 07 Jan 1997

Published: 07 Feb 1997
Electronic identifier: 1359-0278-002-00109

Folding & Design 07 Feb 1997, 2:109–114

© Current Biology Ltd ISSN 1359-0278

Introduction

Kinetics is one of the most exciting areas in current protein folding studies, because a good model for equilibrium properties is believed to have been found by examining the freezing of heteropolymers and borrowing concepts from the physics of spin glasses [1–3]. Indeed, the random energy model (REM) [4] is the most common way to understand heteropolymer freezing. One main strength of REM is its simplicity, both in terms of mathematical tractability and the physical picture it infers. The hallmark assumption of REM is that energies of states are statistically independent over disorder. If this assumption were exact, then there could be no kinetic bias toward the native state and thus the Levinthal paradox of searching through a cosmological number of conformations could not be resolved [5]. Therefore, the problem of heteropolymer folding kinetics is deeply connected to the restricted applicability of REM. Indeed, the applicability of REM is very delicate and nontrivial [6]. REM obviously cannot be exact. For example, because two similar conformations (i.e. large overlap) will have similar energies, the energies of these conformations will be statistically dependent. Thus, the validity of REM involves aspects of the nature of conformations. For dense conformations, the possibilities of rearrangements on small scales, and thus pairs of similar conformations, are rare; this leads to a first order freezing transition and the similar, statistically dependent states do not play an important role in the thermodynamics. However, the states that are similar to the ground state are crucial for kinetics, as they form a path to the native state.

It was expected from the very beginning that kinetics would strongly depend on the character of the heteropolymer sequence. In particular, it is obvious that kinetics would be different depending on how low the native state energy, E_{NS}, is to which (or from which) the kinetic process goes. It is also known that E_{NS} is an important parameter to understand freezing, as it describes the degree of optimization ('minimal frustration') of the sequences.

What is difficult about folding kinetics is the enormously complex topology of the relevant conformation space. To circumvent this problem, we shall assume that kinetics can be viewed in terms of a single reaction coordinate, namely the overlap Q with the native state conformation [7,8]. Taking Q as the only relevant coordinate for kinetics is indeed a delicate and nontrivial approximation. Clearly, we can always choose any coordinate to follow the kinetics, and Q is suitable at least in the sense that it does change when the system travels between folded and unfolded states. Moreover, Q has the important advantage that local physical movements of a real protein chain as well as elementary moves of the Monte Carlo dynamics for a lattice model correspond to small (local) displacements along Q [8]. However, we must still determine the relevant landscape associated with this coordinate. In our approximation, we assume that all other degrees of freedom relax much faster than Q. If that is the case, at each Q our system does have time to equilibrate all other degrees of freedom. Thus, the system can be characterized with the free energy $F(Q)$, which represents the partial thermodynamic

equilibrium of the system, where Q is supposed to be fixed (or quenched) while all other degrees of freedom are supposed to be in equilibrium (or annealed).

This is very similar to what one normally assumes in chemical kinetics, in the theory of transition state [9]. If this is a valid assumption, then we can describe our kinetics as the diffusive motion along the Q coordinate in the $F(Q)$ profile. If, however, the equilibration along the Q = constant hyperplane in the phase space is just as slow as that along Q, then our assumption is totally wrong and the $F(Q)$ profile does not provide us with any useful information about the relevant landscape. Indeed, for long chains, there is no direct connection between two states that have the same Q values due to the same amounts of 'native' bonds, but located in, say, the opposite ends of the globule; the transition between two such states will certainly be slower than that between each of those states and other close states with different Q. Thus, for very long chains, the Q coordinate will certainly not be the slowest one; a theory for this case remains to be found. Our recent theoretical studies indicate that considering Q as a reaction coordinate might in fact be a relatively poor approximation in the quantitative sense even for chains as short as 18-mers. Nevertheless, the theory that employs Q in the capacity of the reaction coordinate appears to be instructive and illuminating in many respects, because it shows the connections between REM breaking down and the folding kinetics. This is why we shall assume in this paper that Q is indeed an appropriate reaction coordinate in the exact and full sense of these words. However, with this choice of reaction coordinate, the problem is reduced to the calculation of the free energy $F(Q)$, where Q is the overlap with the native state (denoted by NS). The $F(Q)$ profile should be calculated as dependent on the sequence, i.e. on how low the ground state energy is for the given sequence. We will perform this calculation for the most general Hamiltonian of short-range pairwise interactions without any assumptions as to the nature of energy correlations.

Results

The model and its free energy profile

We repeat that in order to relate the thermodynamics to the kinetics of protein folding, our main assumption is that Q is the only relevant reaction coordinate. Thus, we have to consider the free energy, $F(Q)$, of the polymer, which is fixed, to have an overlap Q with its native state conformation. To this end, we employ the standard Hamiltonian of short-range pairwise interactions [3]:

$$H = \sum_{I \neq J} B(s_I, s_J)\Delta(\mathrm{r}_I - \mathrm{r}_J) \tag{1}$$

where $s_I \in \{1\ldots,n\}$ is the species of monomer number I, n is the number of species, B_{ij} is the matrix of species interactions, r_I is the position of monomer I, and $\Delta(\mathrm{r} - \mathrm{r}')$ is unity for nearest neighbors and vanishes otherwise. We implicitly assume here that position vectors r_I are such that the conditions of chain connectivity, excluded volume, and constant density are all met. With the Hamiltonian (eq. 1), the free energy of interest can be written in a general form, as in equation 2:

$$F_{seq}(Q) = -T \ln\left[\sum_C e^{-H_C/T}\Delta(Q_{C,\mathrm{NS}} - Q)\right] \tag{2}$$

where the subscript *seq* indicates that the free energy is written for a given sequence, summation is performed over conformations C, and $Q_{C,\mathrm{NS}}$ is the overlap (the number of bonds in common) between conformations C and NS: $Q_{C,\mathrm{NS}} = \Sigma_{I \neq J}\Delta(\mathrm{r}_I^C - \mathrm{r}_J^C)\,\Delta(\mathrm{r}_I^{\mathrm{NS}} - \mathrm{r}_J^{\mathrm{NS}})$. The dependence on the sequence enters both through the Hamiltonian and through the ground state conformation for the given sequence. As we expect freezing to occur at temperatures higher than that of the phase segregation of monomers, we can use a high temperature expansion to describe freezing [3]. This yields equation 3:

$$F_{seq}(Q) \simeq \overline{(H)\big|_Q} + \frac{1}{2T}\left[\overline{(H^2)\big|_Q} - \left(\overline{(H)\big|_Q}\right)^2\right] - T \ln M \tag{3}$$

where $\overline{(\ldots)|_Q}$ means the average over all conformations with the given overlap to the ground state, M is the total number of conformations. We have kept terms to $\mathcal{O}(1/T^2)$, which is exact for the independent interaction model (in which the elements of B_{ij} are taken from a Gaussian distribution) [2]; higher order terms may be necessary to quantitatively model a particular set of interactions B_{ij}.

As one can analytically perform only a statistical analysis, we have to average over the sequences. In principle, one could directly average over the ensemble of sequences with the given value of the ground state energy. This ensemble of sequences is in a way similar to the microcanonical ensemble in regular statistical mechanics. It is, however, technically simpler to use the analog of the canonical ensemble, where the native state energy E_{NS} is not fixed, but rather controlled through a temperature T_p (equivalently, $1/T_p$ is a Lagrange multiplier to constrain E_{NS}), and thus it has a Gibbs probability distribution:

$$P_{seq} \sim \exp\left[-E_{\mathrm{NS}}(seq)/T_p\right] \simeq 1 - \frac{1}{T_p}E_{\mathrm{NS}}(seq) \tag{4}$$

where in the last transformation we have resorted to high T_p expansion (which is justified for the same reason as high temperature expansion above). Thus, we characterize a given ensemble of sequences by the value of T_p: for lower T_p, we model sequences whose native states are better optimized energetically. This prescription for energetic optimization of the native state has many incarnations, including 'minimal frustration' [1], 'sequence selection' [10], and 'imprinting' [3]. When one averages the free energy (eq. 3) over the probability distribution (eq. 4), the

result appears very simple, because, due to the structure of the Hamiltonian, the energy depends directly on Q_{NS}, and thus it is very easy to implement the condition of the fixed Q_{NS}, and the result can be written in the form:

$$F(Q) \equiv \sum_{seq} P_{seq} F_{seq}(Q) = E(Q) - TS(Q) \quad (5)$$

where the energy $E(Q)$ is given by:

$$E(Q) \approx -\frac{\tilde{B}^2}{T_p} Q - \frac{\tilde{B}^2}{T}(Q_m - Q) \quad (6)$$

$\tilde{B}^2 \equiv \Sigma_{ij} p_i p_j (B_{ij} - \Sigma_{kl} p_k p_l B_{kl})$ is the variance of the elements of the interaction matrix [3], p_i is the fraction of monomers of type i, Q_m is the maximum overlap, and the entropy $S(Q)$ of conformations with Q bonds in common with NS is:

$$S(Q) = \ln M - Q\left(s - \frac{3}{2}\ln\frac{Q}{Q_m}\right) - 2Q_m\left(\frac{Q}{Q_m} - \frac{1}{2}\right)^2 \quad (7)$$

For simplicity, we have made the assumption that all conformations have the same number of total contacts Q_m and differ only by the number of native contacts Q (which is relevant for heteropolymers in a poor solvent). One can also consider the regime in which only native contacts form (which is relevant for well designed heteropolymers in a theta solvent); while this second regime changes some of the equations below, there appear to be no qualitative differences for the issues we will address.

The first energy term comes from the mean of the density of states $\langle \bar{H} \rangle$ and describes how the energy is pulled down due to selection; indeed, to this order, the ground state energy is $E_{NS} = -Q_m \tilde{B}^2/T_p$, and thus this term says that on average, the energy is given by the energy of the native state times the fraction of native contacts.

The second energy term comes from the width of the density of states, resulting from the $\langle \bar{H}^2 \rangle - \langle \bar{H}^2 \rangle$ terms; the Q dependence of this width enters from the correlator $\langle \bar{H}^2 \rangle$ [6], as two conformations with a given overlap Q with NS also have Q bonds in common, on average. Higher order terms can modify the Q dependence of the width, depending on the nature of interactions. The Q dependence of the mean and width of the density of states is a correction to REM, describing the nature of energy correlations. Simply put, the energy terms in equation 6 say that each native bond typically has energy $-\tilde{B}^2/T_p = E_{NS}/Q_m$ and each nonnative bond contributes $-\tilde{B}^2/T$.

The first two entropy terms detail how many conformations exist for a fixed set of Q contacts; as each common contact pins two parts of the chain together, for contacts Q, the number of conformations is the loop entropy [11–13] of $\mathcal{O}(Q)$ loops of length $\mathcal{O}(1/Q)$. The third entropy term results from how many ways one can choose Q bonds from the total Q_m (using Sterling's formula) [13]. One can easily write and explore a more general expression, where Sterling's formula is not used. We choose not to do it here, as the expression (eq. 7) is only a crude approximation anyway (because we consider all pairings to be equivalent) and thus this expression purports to be only a qualitative example.

Typical conformational spaces can be characterized by a region with an exponential number of states $Q < Q_d$ compared with a discrete region $Q > Q_d$, where the probability of finding a pair of conformations with $P(Q_d \le Q < Q_m) < P(Q_m)$, where $P(Q) = \Sigma_{C1,C2}\, \Delta(Q - Q_{C1,C2})$ is the probability that two conformations overlap Q. Thus, Q_d is a purely geometrical characteristic of a given conformation space. While there are certainly many aspects in which conformation spaces differ, from the standpoint of freezing, Q_d is important as it indicates when the conformation space 'runs out' of conformations as we approach greater overlap with a desired conformation. This idea has also been described in a slightly different fashion elsewhere [8,14].

Thus, we choose s such that $S(Q = Q_d)$ vanishes. For lattice models, we can measure Q_d (see Fig. 1 in [6]). For proteins, the measurement of Q_d is less clear, but perhaps one may use arguments that compare topological aspects of proteins and lattice models in order to draw comparisons; for example, it has been hypothesized that many proteins are topologically equivalent to 27-mers [14].

Thus, we find a free energy which has two minima at $Q = 0$ and Q_m. The relative depths of these minima depend on T and T_p, with a first order freezing transition occurring when $F(0) = F(Q_m)$. As energy correlations affect only the region between these free energy minima, the result for the freezing temperature is exactly what was found for REM. Indeed, as all aspects of the phase diagram in T and T_p depend only on the $Q = 0$ and $Q = Q_m$ cases, the REM corrections do not change the REM result (see phase diagrams in [3]). However, the kinetics will depend on the free energy for intermediate Q, as the chain must go through these states during folding.

Kinetics

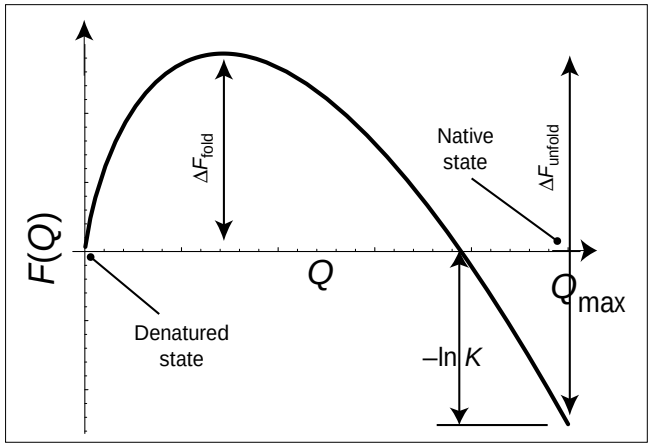
To examine the nature of the kinetics predicted by our model, we calculate the height of the free energy barrier (Fig. 1). For the free energy (eq. 5), we find a maximum at $Q = q_{TS} Q_m$, determined self consistently by equation 8:

$$q_{TS} = \exp\left[\frac{8}{3} q_{TS} - \frac{7}{3} + \frac{2}{3}\ln K\right] \quad (8)$$

where

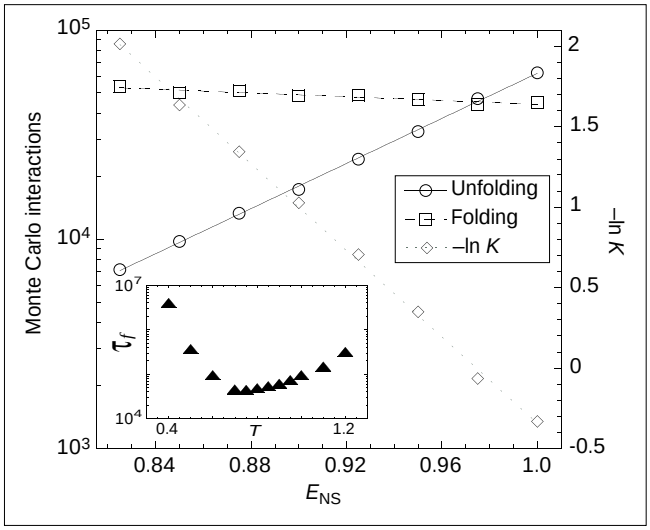
$$\ln K \equiv \frac{F(Q_m) - F(0)}{Q_m T} = s + \frac{\tilde{B}^2}{T^2} + \frac{E_{NS}}{Q_m T} \quad (9)$$

Figure 1



Typical free energy plotted against overlap with the target conformation.

Figure 2



Folding and unfolding time plotted against E_{NS}. Inset: temperature dependence of folding time τ_f.

is the free energy difference between native and denatured states [3] and we have substituted $E_{NS} = -Q_m\tilde{B}^2/T_p$. We see that q_{TS} depends solely on K.

Inserting equation 8 for Q_{TS} into equation 5, we find the barrier heights for folding and unfolding respectively:

$$\frac{F_f}{Q_m T} = \frac{3}{2}q_{TS} - 2q_{TS}^2$$

$$\frac{F_u}{Q_m T} = \frac{3}{2}q_{TS} - 2q_{TS}^2 - \ln K \qquad (10)$$

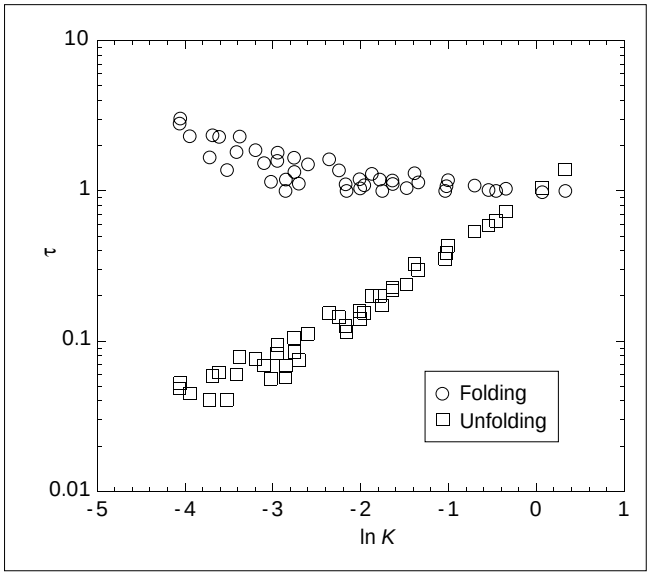
It is surprising that kinetics can be characterized in terms of a single parameter K, which has the physical interpretation of the thermodynamic stability of the native state.

We now compare the remaining aspects of the theory with Monte Carlo kinetics simulations. As our goal was to make a simple model in order to extract some 'robust' physical property of folding, we do not expect an exact quantitative agreement with simulation. However, there are several predictions we can test. We will employ 18-mers as a model system because we have demonstrated the validity of our approximation for 18-mers, they fold quickly, and we have the possibility of numerically comparing the exact thermodynamics with the kinetics, as we are doing analytically. The general character of 18-mer kinetics is shown in Figure 2, in which we have plotted the folding time, unfolding time, and K versus the degree of optimization of the native state. (To create sequences with arbitrary ground state energies, we use a trick: we fix the sequence and vary the interaction matrix. We sequence annealed [10] to a perfectly optimized sequence for a given IIM interaction matrix B_{ij}^d and then created a second, Gaussian random IIM matrix B_{ij}^f [both matrices have zero mean and unit variance]. To obtain arbitrarily optimized sequences for a given matrix, one can vary a parameter $g \in [0,1]$ to define a new matrix $B_{ij} = gB_{ij}^p + (1-g)B_{ij}^d$. For $g = 1$, we obtain the original matrix, and the full optimization. As one decreases, we decrease the optimization of the native state up to the point where it is no longer the ground state [for 18-mers, this occurred at $g \approx 0.825$].) For the folding time, we have taken the mean first passage time, and for unfolding, we have taken the mean first passage time to go from the native state to a denatured state with $q_{NS} = 0.375$. As the native state is less optimized, folding time τ_f increases exponentially and unfolding time τ_u decreases exponentially. Also note that there is good agreement with the theory, as K measured by $\ln K = (F_f - F_u)/T = \ln \tau_f/\tau_u$ has a near perfect linear relation with E_{NS}. Also, the temperature dependence of 18-mer folding is similar to that found in other studies [7,8,15].

A much more striking agreement is obtained when we compare the folding and unfolding times with the measured value of K ('Brønsted plot' [16]). We examined eight temperature values from $T = 0.65\tilde{B}$ to $T = 1.0\tilde{B}$ and eight values of E_{NS} ranging from 82.5% to 100% of the maximal possible optimization and performed 250 folding and unfolding runs for each (T, E_{NS}) coordinate. To normalize out effects due to the temperature-dependent diffusion constant [8], we divided the τ_f and τ_u by the folding time for $E_{NS} = E_{NS}^{max}$ at each temperature. Figure 3 shows that and τ_f and τ_u data collapse very well, as predicted by our theory. Also, the measured value of K from non-equilibrium kinetics agrees well with that obtained from the exact enumeration; however, in order to match to the ana-

lytic prediction (eq. 9), we had to use adjusted values for E_{NS}. This is not surprising, as our theory does not purport to be quantitative, but instead tries to find the universal aspects of folding of short chains. Moreover, our theory appears to be self consistent and to capture the essence of the problem: K appears to be a single parameter to characterize kinetics. In order to improve the agreement of analytic treatment and lattice models, perhaps better modeling of the polymer entropy is necessary [13]. Also, an unresolved aspect of Monte Carlo simulations is whether the transition state consists of a unique nucleus of native contacts. Interestingly, simulations show that this depends on the nature of conformational space: simulations of 27-mers yield many nuclei whereas 36-mers and 64-mers indicate a single nucleus [7,17]. Our theory suggests that if $Q_{TS} \geq Q_d$, then the entropic possibilities for the transition state are so low that only a unique nucleus may be found. This can occur either by a small Q_d (e.g. Q_d appears to decrease as chain length increases) or sufficiently large Q_{TS} (e.g. Q_{TS} increases with temperature and with lessening the degree of energetic optimization of the native state). The validity of this prediction, which can be tested by lattice simulations for example, will be an excellent test of our theory.

Figure 3



Collapse of data with a variety of T and E_{NS}.

Discussion

We have constructed a simple, nonphenomenological, and non-REM model of the free energy profile $F(Q)$ for proteins and have used this to describe the folding and unfolding kinetics. While our model qualitatively agrees with computer simulations, corrections to the thermodynamics of our model, including higher order terms from the high temperature expansion as well as the incorporation of the density fluctuations and the possibility of globule-to-coil decollapse, may be necessary for stronger quantitative agreement. However, the goal of this work was to find some robust, perhaps even 'universal', character of folding kinetics of short proteins. Indeed, the universal Brønsted plots of folding and unfolding rates versus K are potentially such a characteristic, and recent experiments seem to support this conclusion [16,18]. Also, our model predicts Hammond behavior: Q_{TS} increases (moves toward the NS) as K increases (NS is less stable). Experiments indicate that while the transition state indicates Hammond behavior overall, particular elements (such as helix1 of barnase) can actually have anti-Hammond behavior [19]; this is a good example of what is beyond the reach of our theory, as we cannot consider, for example, the nature of secondary structure.

Moreover, we do not expect that this approximation is universally applicable; indeed, the folding time scales exponentially with N in this model (in $N \to \infty$ limit), implying that sufficiently long chains will have prohibitively long folding times and therefore fold by a potentially different mechanism. While one may still expect diffusion-like motion in Q for longer chains, the system may not reach quasi-equilibrium for a given Q during dynamics, and thus one could not use $F(Q)$ as an effective potential to describe kinetics.

Also, one may need to consider other order parameters, such as the total number of contacts: one could have two conformations with the same number of native contacts, but that differ greatly in the number of nonnative contacts, which should make a qualitative difference in folding simulations that begin from each of these conformations. This problem has perhaps been circumvented here, as the sequences are well designed and the mean interaction between monomer species is zero; thus, few nonnative bonds form. Finally, one would expect, purely due to entropic considerations, that not all native bonds are equally important in folding; these differences are not incorporated in a model that takes the total number of native contacts as a reaction coordinate as it treats all bonds equally. An analysis of these entropic factors is beyond the scope of this paper and is an important unresolved issue.

Acknowledgements

We have benefited from discussions with JN Onuchic and EI Shakhnovich. The work was supported by NSF (DMR 90-22933 and MCB-93-16186). AY Grosberg acknowledges the support of a Kao Fellowship.

References

1. Bryngelson, J.D. & Wolynes, P.G. (1987). Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA* **84**, 7524–7528.
2. Shakhnovich, E. & Gutin, A. (1989). Formation of unique structure in polypeptide chains: theoretical investigation with the aid of a replica approach. *Biophys. Chem.* **34**, 187–199.
3. Pande, V.S., Grosberg, A.Y & Tanaka, T. (1995). How accurate must

potentials be for successful modeling of protein folding? *J. Chem. Phys.* 103, 9482–9491.
4. Derrida, B. (1980). Random-energy model: limit of a family of disordered models. *Phys. Rev. Lett.* 45, 79–82.
5. Bryngelson, J.D. & Wolynes, P.G. (1989). Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.* 93, 6902–6915.
6. Pande, V.S., Grosberg, A.Y., Joerg, C. & Tanaka, T. (1996). Is heteropolymer freezing well described by random energy model? *Phys. Rev. Lett.* 76, 3987–3990.
7. Šali, A., Shakhnovich, E. & Karplus, M. (1994). How does a protein fold? *Nature* 369, 248–251.
8. Socci, N.D., Onuchic, J.N. & Wolynes, P.G. (1996). Diffusive dynamics of the reaction coordinate of protein folding funnels. *J. Chem. Phys.* 104, 5860–5868.
9. Eyring, H. & Eyring, E.M. (1963). *Modern Chemical Kinetics.* Reinhold, New York.
10. Shakhnovich, E.I. & Gutin, A.M. (1993). 'Evolutionary' selection of protein sequences: a new insight on protein design. *Proc. Natl. Acad. Sci. USA* 90, 7195–7199.
11. Grosberg, A.Y. & Khokhlov, A.R. (1994). *Statistical Physics of Macromolecules.* AIP, New York.
12. Gutin, A.M. & Shakhnovich, E.I. (1994). Statistical mechanics of polymers with distance constraints. *J. Chem. Phys.* 100, 5290–5293.
13. Plotkin, S.S., Wang, J. & Wolynes, P.G. (1996). Correlated energy landscape model for finite, random heteropolymers *Phys. Rev. E* 53, 6271–6274.
14. Onuchic, J.N., Wolynes, P.G., Luthey-Schulten, Z. & Socci, N.D. (1995). Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. USA* 92, 3626–3630.
15. Socci, N.D. & Onuchic, J.N. (1995). Kinetic and thermodynamic analysis of protein-like heteropolymers: Monte Carlo histogram technique. *J. Chem. Phys.* 103, 4732–4744.
16. Fersht, A.R., Itzhaki, L.S., ElMasry, N.F., Matthews, J.M. & Otzen, D.E. (1994). Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc. Natl. Acad. Sci. USA* 91, 10426–10429.
17. Abkevich, V.I., Gutin, A.M. & Shakhnovich, E.I. (1994). Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry* 33, 10026–10036.
18. Milla, M.E., Brown, B.M., Waldburger, C.D. & Sauer, R.T. (1995). P22 Arc repressor: transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. *Biochemistry* 34, 13914–13919.
19. Matthews, J.M. & Fersht, A.R. (1995). Exploring the energy surface of protein folding by structure–reactivity relationships and engineered proteins. *Biochemistry* 34, 6805–6814.

Because *Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/fad.htm – for further information, see the explanation on the contents page.